

Questions to validate your knowledge

1 Variables and data

1. What is a data file?
2. Explain the ways of storing data in a plain form and as values and frequencies.
3. Having data 8 6 2 8 9 7 5 4 - determine their ranks.
4. Which characteristic of data file express their level and which ones describe their variability?
5. Determine median of the data file 2 4 6 1 7 3 9 .
6. Compute average and variance of data file 3.15 2.22 10.45 3.57 1.58 6.9 5.13 -0.75 2.24 6.58 4.65 6.52 3.79 4.95 1.87 3.87 4.22 3.68 6.3 3.64 ?
7. What is the mode of data file 2 5 4 6 4 2 4 2 6 2 2 2 6 4 2?
8. What is the difference between general bar graph and histogram?

2 Probability and random variable

1. What is random experiment?
2. What is an event?
3. What data produces the experiment of flipping a coin?
4. What data produces random variable describing the experiment of flipping a coin?
5. Which are the three important properties of the probability?
6. Define classical definition of probability.
7. Define statistical definition of probability.
8. What is the major difference between classical and statistical definition of probability.
9. After ten flips of coin we obtained 3 heads and 7 tails. We concluded that the probability of head is 0.3. Which definition of probability we have used?
10. We inspected a coin and concluded that it is not damaged. So we determined the probability of head is 0.5. Which definition of probability we have used?

11. What is the definition of conditional probability?
12. Consider an experiment of tossing a dice. What is the probability of even number if we know, that the number that really fell is less than 4.
Hint: Choose only from those that could have fallen.
13. Consider an experiment of drawing colored balls from a box. The drawn ball is not returned back. Are the draws independent?
Hint: Is the probability of drawing specified color all the same during the draws?
14. We have an experiment of flipping two coins. The results are: “both heads”, “both tails” and “different sides”. Are the probabilities of these result equal?
15. A natural definition of independence x and y is $P(x|y) = P(x)$ - the knowledge of y does not influence the probability of x . Using the definition of conditional probability, derive the formula $f(x, y) = f(x) f(y)$.

3 Description of random variable and vector

1. What is the difference between random experiment and random variable.
2. What types of random variable do you know?
3. Can random variable have negative values?
4. What is the merit of random vector in comparison with random variable?
5. What are realizations of a random vector?
6. Define distribution function.
7. What are the basic properties of a distribution function?
8. Determine the probability $P(a, b)$, $b > a$ using the distribution function.
9. What is the probability $P(X = 5)$ for a continuous random variable X equal to?
10. Is any distribution a continuous function?
11. Define probability function of discrete random variable.
12. Define density function of continuous random variable.
13. What are the basic properties of any probability or density function.
14. What assertion is correct: all values of any probability function are (i) non-zero, (ii) positive, (iii) non-negative.

15. What assertion is correct: all values of any density function are (i) non-zero, (ii) positive, (iii) non-negative.
16. How the probability $P(a, b)$, $b > a$ can be computed using probability or density function?
17. What is the definition of expectation for discrete random variable?
18. What is the definition of expectation for continuous random variable?
19. Define α -quantile and α -critical value of continuous random variable.
20. How can you find median of continuous random variable.
21. How can you find mode of continuous random variable.
22. We have a function $y = kx$. Determine the constant k so that this function would be a density function of the interval $x \in (0, 5)$.
23. The probability function of random variable X is given by the table

x	1	2	3	3	5
$f(x)$	0.2	0.1	0.3	0.2	0.2

Determine the distribution function.

24. Write the probability functions describing the experiments (i) “flipping a fair coin” and (ii) “flipping some unfair coin”.
25. The random variable X describes the following experiment “time of waiting for a bus with an interval 5 min if your coming to the bus station is random”. Write the density and distribution functions of this random variable.
26. We have two random variables X and Y with probability functions $f(x) = 2 - 2x$ and $f(y) = 2y$, both on $x \in (0, 1)$ and zero otherwise. Write the joint density function $f(x, y)$.
27. A random vector $[X, Y]$ has joint density function

$$f(x, y) = k \exp(x^2 + 2y^2)$$

for $x, y \in (-\infty, \infty)$. k in a normalization constant. Are these random variables independent?

28. A random vector $[X, Y]$ has joint density function $f(x, y) = 1$ on $x \in (0, 1)$ and $y \in (0, 1)$. Determine the probability $P([X, Y] \in (0, 0.1) \times (0, .2))$.
29. A random vector $[X, Y]$ has joint probability function given by the table

x/y	1	2	3
1	0.1	0.05	0.3
2	0.35	0.1	k

Determine the value of k .

30. A random vector $[X, Y]$ has joint probability function given by the table

x/y	1	2	3
1	0.2	0.1	0.1
2	0.3	0.1	0.2

Determine the marginal $f(y)$ and the conditional $f(x|y)$.

31. A random variable X has probability function $f(x) = p^x(1-p)^{1-x}$, $x = 0, 1$ and $p \in (0, 1)$.
- What is the probability that in the next sampled value of X will be 1?
 - What is the expected number of results $x = 1$ in 100 experiments?

4 Important distributions

- What is the definition of Bernoulli distribution?
- What is the meaning of the parameter p in Bernoulli distribution?
- The probability that a newborn will be a boy is 0.52. What is the probability that in a family with 5 children there will be 2 boys and three girls?
- Describe the experiment connected with a binomial random variable.
- For several years we have measured accidents in five points of a large traffic region. The results were

point of measurement	1	2	3	4	5
numb. of accidents	38	147	51	223	197

On the basis of the measured data determine the probability function describing these accidents.

- We are flipping a fair coin. What is the probability, that the “head” appears at the third flip for the first time.
- One percent of bits transmitted through a digital transmission are received in error. Bits are transmitted until the first error. Let X denote the number of bits transmitted until the first error. What distribution describes random variable X ?

8. Compare the supports of Normal distributions $N(0, 1)$ and $N(10, 3)$, where the denotation is $N(\mu, \sigma^2)$.
9. A fixed distance 10 meters is repeatedly measured. We define two random variables X - the value of the measurement and Y - the error of measurement from the true value. What is the distribution of these random variable and what they differ in?
10. What are the main assumptions of a uniform distribution.
11. We have a random variable with uniform distribution on the interval $(3, 9)$. What is the probability that a realization of X will be within the interval $(5, 7)$.

5 Regression analysis

1. Write regression line constructed only for two measured points: $x_1 = [1, 1]$, $x_2 = [2, 5]$. What will be the correlation coefficient?
2. The regression line is computed for three data points

$$x_1 = [1, 2], x_2 = [2, 2], x_3 = [3, 5]$$

Compute the criterion as a sum of squares of residuals.

3. We investigate a profit (y) in dependence on invested money (x). We obtained linear regression $y = 0.21x - 100$. What will be the profit if we invest 5000 (Kč)?
4. We investigate a profit (y) in dependence on invested money (x). We obtained linear regression $y = 0.21x - 100$. How much we need to invest to have the profit equal to 1000 (Kč)?
5. We have exponential regression $y = b_0 \exp(b_1 x)$. Perform its linearization.
6. Write a quadratic and cubic regression curves.
7. The polynomial regression has coefficients $b_0 = 2.1$, $b_1 = 0.6$, $b_2 = 1.2$ and $b_3 = 0.1$. Write the value of the prediction at $x = 2$.

6 Population and data sample

1. What are the differences between population and random sample?
2. What is the difference between random sample and sample realization?
3. Is an average of random sample a number or random variable?

4. Is an average of sample realization a number or random variable?
5. We measure speeds of passing cars. What is the population, what random sample and what sample realization?
6. We throw ten times a dice. What is the population, what random sample and what sample realization?
7. We monitor speeds of cars on a certain point of a motorway. We assume that the speeds are normally distributed with the variance 5.83. What parameter we need to estimate?
8. What is the difference between parameter and its point estimate?
9. What is the statistics in stochastic estimation?
10. What are the most important characteristics of random sample?
11. What are the expectation and variance of sample average equal to?
12. What is the meaning of the formula expressing the expectation of sample average?
13. What is the meaning of the formula expressing the variance of sample average?
14. Which properties has the sample average.
15. When comparing efficiency of two sample averages with different lengths of sample, which one would be better (has higher efficiency)?

7 Statistical inference

1. What statistics is suitable for estimation of expectation?
2. What statistics is suitable for estimation of variance?
3. What statistics is suitable for estimation of proportion?
4. What statistics is suitable for testing of independence?
5. A sample of the length n is taken from normal population with expectation μ and variance σ^2 . What is the distribution of sample average?
6. We have a population $f(x)$. We want to compute a confidence interval for its unknown parameter μ . We chose the statistics \bar{x} (sample average). Which distribution is used: $f(x)$ or $f(\bar{x})$?
7. What is the difference between both-sided, left-sided and right-sided interval?
8. What are zero and alternative hypotheses?

9. What are region of acceptance and critical region?
10. What is realized statistics T_r ?
11. Explain the difference between both-sided, left-sided and right-sided test.
12. What is the connection between confidence interval and test of hypothesis.
13. What is the conclusion of a test if the realized statistics lies in the critical region?
14. What is the conclusion of a test if the p -value is smaller then the confidence level?

8 Important tests

1. What is the difference between parametric and nonparametric test.
2. Which are the parametric tests with one sample?
3. Which are the parametric tests with two samples?
4. Which are the parametric tests with more samples?
5. Which are the nonparametric tests with one sample?
6. Which are the nonparametric tests with two samples?
7. Which are the nonparametric tests with more samples?
8. Which samples of independence do you know?
9. Which are the tests of a distribution?

12 Validation in regression analysis

1. How can you evaluate the quality of regression analysis according to the xy -graph?
2. What is the statistics for Pearson t -test?
3. Is a regression analysis suitable if the p -value of the Pearson test is very small?
4. Is a regression analysis suitable if the p -value of the F -test is very small?
5. Is a regression analysis suitable if the sum of squares of residuals is very small?
6. Should residuals be independent or dependent?